Résumé

<u>Meilleures pratiques en matière de méthodes de recherche</u>

Facteurs à prendre en considération dans le choix d'une méthode de collecte électronique de données

Karen L. Courtney et Catherine K. Craven

Les chercheurs et les chercheuses sont de plus en plus nombreux à envisager l'adoption de méthodes de collecte électronique de données — lesquelles comportent l'utilisation d'Internet, du courrier électronique ainsi que d'ordinateurs de bureau et de poche — pour les enquêtes, les carnets d'écoute, les instruments de recherche et les groupes de discussion. Se fondant sur les résultats de recherches antérieures sur la collecte électronique de données, les auteures présentent les facteurs clés, qui comprennent les caractéristiques liées à la population, aux données et aux ressources, qui guident les chercheurs et les chercheuses dans le choix de la méthode de collecte de données la plus appropriée à la question de leur recherche. Les facteurs dont il faut tenir compte dans les décisions relatives à la collecte de données sur la population sont l'âge, le sexe, le statut socioéconomique et l'accès aux technologies. Les facteurs clés qui influencent le choix d'une méthode de collecte de données comprennent la sensibilité du sujet, l'importance du facteur temps, les données longitudinales et les données contextuelles. Les décisions relatives à la collecte de données sont aussi fondées sur les ressources en termes d'argent, de temps et de technologies. Les technologies et la démographie vont continuer de modifier les méthodes et les possibilités de collecte de données ainsi que d'influer sur elles. Cependant, par l'examen de ces facteurs clés, les chercheurs et les chercheuses pourront prendre les décisions en matière de collecte de données qui sont appropriées à chaque projet.

Mots clés : Internet, collecte de données, ordinateurs de poche, méthodologie de recherche, informatique médicale

Best Practices in Research Methods

Factors to Weigh When Considering Electronic Data Collection

Karen L. Courtney and Catherine K. Craven

Researchers are increasingly considering the adoption of electronic data-collection methods — which entail the use of the Web, e-mail, and desktop and handheld computers — for surveys, diaries, research instruments, and focus groups. Based on prior research findings on electronic data collection, the authors delineate the key factors, which incorporate population, data, and resource characteristics, that guide researchers in selecting the data-collection method most appropriate for their research question. Population factors to consider in data-collection decisions are age, gender, socio-economic status, and access to technology. Key data factors influencing the selection of collection method include sensitivity of the topic, time sensitivity, longitudinal data, and contextual data. Data-collection decisions are also based on financial, time, and technological resources. Technology and demographics will continue to change and affect data-collection methods and possibilities. By examining these key factors, however, researchers will be able to reach data-collection decisions that are appropriate for each project.

Keywords: Internet; data collection; computers, handheld; research design; medical informatics

In the United States, 63% of adults participate in Internet activities such as e-mail exchange, Web browsing, instant messaging, and participation in chat rooms (Madden, 2003). As the number of online participants grows, more researchers are considering the use of electronic data collection (EDC) methods. With EDC methods, the Web, e-mail, and desktop and handheld computers are used for completing surveys and diaries, taking part in focus groups, and participating in other aspects of research. The literature contains numerous articles that partly explore the challenges and opportunities presented by EDC. We found that most health research entailing EDC methods touches on one or more of thirteen issues, three of which are time, cost, and immediacy (Courtney & Craven, 2004). There is a dearth of resources, however, to guide researchers in making decisions about EDC methods. This paper describes factors that can guide investigators in selecting the most appropriate data-collection method for their research question (Figure 1).

Figure 1 *Data–Collection Method: Factors to Consider*
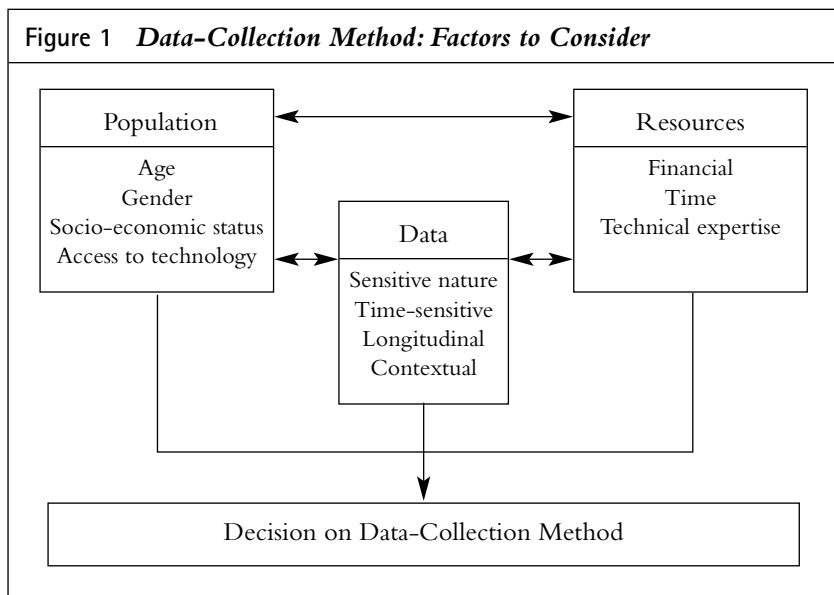
These critical decision-making factors, identified through an extensive literature review, include population, data, and resource characteristics (Courtney & Craven).

To identify articles on EDC, in spring 2004 we searched the databases MEDLINE, CINAHL, Compendex, and Dissertation Abstracts. Our article-retrieval search strategy comprised keywords, controlled terminology terms, and hand searches. Of the 205 articles identified, 71 entailed the use of EDC in research studies. We reviewed those 71 publications and found that 43 reported evidence-supported findings, 14 were review or concept articles, and 14 were technical or design descriptions.[1] The findings are presented in relation to each factor (i.e., population, data, and resource characteristics) that should be considered in decision-making on data-collection methods.

Population

To determine whether an EDC method is suitable for a proposed project, the researcher must take into account the characteristics of the study population. In the United States there is a potential for sampling bias when EDC methods are used, because only 63% of the population is online (Madden, 2003). Researchers must evaluate access to technology

[1] A full bibliography is available upon request.

in relation to the following key population factors: age, gender, socio-economic status, and race and ethnicity.

Age

Age is an important consideration for methods involving use of e-mail and the Internet, because the majority of those online are young (Greenspan, 2003). Of Internet users in the United States, 61 million are between the ages of 25 and 49 (Greenspan); 87% of youths between the ages of 12 and 17 are Internet users (Lenhart, 2005). Researchers, however, must revisit this generalization regularly, because the demographics of Internet users are changing. One of the fastest-growing segments of Internet users is seniors (those aged 65 and over). Between 2000 and 2004 the number of Internet-using seniors increased by 47% (Fox, 2004); eight million seniors, or 22%, are currently online (Fox). In fact, projections are that 65% of adults aged 50 to 65 and 37% of those over 65 will be online by 2006, so in the future these populations might be more accessible via EDC methods (Greenspan). The use of EDC methods has been successfully demonstrated across the lifespan. These methods have been used with children as young as 8 (Palermo, Valenzuela, & Stork, 2004) and adults as old as 75 (Chang, 2004).

Gender

Although gender is a population characteristic worth considering, data suggest it is not necessarily a factor in use of electronic methods involving the Internet. Contrary to perceptions that the typical Internet user is male, female users make up more than half (51%) of all Internet users, and their numbers are increasing (Madden, 2003). Despite the difference in the online activities pursued by each gender, researchers have not noted a gender bias in participant recruitment for studies using EDC (Pryor et al., 2002). In their university-campus survey on a conjunctivitis outbreak, Pryor et al. found that Web-based methodology yielded a more representative sample of the campus population than a paper-based survey.

Socio-economic Status

Decisions on data-collection methodology should also include the socio-economic status of the target population. In terms of Internet-based tools, on a typical day only 37% of US households with an annual income of less than $30,000 are likely to send an e-mail — the most common online activity — compared to 58% of households with a annual income of more than $75,000 (Madden, 2003). A population's level of education also affects online activities: "For example, while 39% of high school graduates were sending e-mail on a typical day in

December 2002, 61% of college graduates were doing so" (Madden, p. 10). Area of residence is another socio-economic factor to consider; only 52% of rural dwellers use the Internet, compared to 66% of suburban and 67% of urban dwellers (Bell, Reddy, & Rainie, 2004).

Race and Ethnicity

Race and ethnicity are potential social factors to consider. Persons with non-white heritage participate in Internet activities less than their white counterparts. In one study, 51% of African-Americans and 62% of English-speaking Hispanics said they were online (Madden, 2003).

Online growth rates are changing for persons with different demographic characteristics. Such factors might be less critical to consider in future decision-making on data-collection methods.

Data

The nature of the data to be collected is another critical concern when choosing a collection method. Data factors that influence this decision include sensitivity of the topic, time sensitivity, longitudinal data, and contextual data.

Sensitivity of the Topic

Prior research indicates that sensitive topics such as intimate-partner violence, substance abuse, psychological state, and sexual behaviour lend themselves to EDC especially well (Buchanan & Smith, 1999; Rhodes, Lauderdale, He, Howes, & Levison, 2002; Turner et al., 1998). For example, researchers have found that electronic screening via a laptop computer can be more effective in identifying potential victims and perpetrators of domestic abuse than face-to-face clinician evaluations (Rhodes et al.). In a study of adolescent male sexual behaviour, illegal substance use, and violence, Turner et al. found substantial differences in highly sensitive data reporting when comparing personal computer-based and traditional paper-based self-administered questionnaires: male adolescents reporting electronically were more likely to report sex with a prostitute (Odds Ratio [OR] 4.24, $p < 0.001$) and specific male-male sexual behaviours (OR ranging from 2.25 to 7.85, $p < 0.05$).

Any discussion of sensitive-topic research must examine the trustworthiness of the data. A number of EDC studies have noted increased sensitive-behaviour reporting, decreased self-revelation, and social desirability bias, which are components of the concept of data trustworthiness (Buchanan & Smith, 1999; Joinson, 1999; Rhodes et al., 2002; Turner et al., 1998). Few studies, however, explicitly acknowledge receiving falsified responses or identities (Cronk & West, 2002; O'Neil, Penrod, &

Bornstein, 2003). EDC methods, subject to some of the same trustworthiness issues as paper-based methods, provide the researcher with additional mechanisms for protecting data integrity. Examples of such safeguards include checking timestamps to verify data-collection time, using cookies in Web-based applications to track the number of responses from one computer, and presenting one question at a time to limit answer changing.

Time Sensitivity

The Internet's relative accessibility to participants makes it a valuable tool for studies that rely on timeliness of data reporting. Likewise, rapid and timely data reporting increases when participants carry data-collection devices with them. As wireless access in a community increases and more people own ever-smaller portable devices such as handheld computers or personal digital assistants (PDAs) and smart phones, more rapid data collection will be possible. For instance, in a conjunctivitis-outbreak study at a university, the speed of electronic methods not only permitted rapid data collection but also allowed clinicians to respond quickly to the local public health crisis (Pryor et al., 2002).

Longitudinal Data

EDC lends itself to longitudinal data collection. In studies of chronic health conditions that require frequent and longitudinal data collection, participants reporting via electronic diaries demonstrated increased compliance. For example, in a study monitoring chronic pain, participants reporting via handheld computers were more compliant in reporting at prescribed intervals, despite a much more stringent definition of compliance for this electronic group, than the paper-based comparison group (Jamison et al., 2001).

Contextual Data

For some research, context is an essential component of data collection. This is especially true for qualitative research. Context is a set of personal, environmental, and socio-cultural factors in which an experience is nested in daily life (Hinds, Chaves, & Cypess, 1992). Collection of contextual data relies on observation and self-reporting. The potential for bias in self-reporting limits the applicability of EDC methods for contextual data without observational confirmation. If a proposed study depends heavily on contextual analysis, non-electronic data-collection methods such as face-to-face interviews or focus groups might be more appropriate. As video phones, Web cameras, and high-speed Internet access proliferate, the potential for the electronic collection of high-quality contextual data will increase.

Resource Characteristics

The availability of resources, always a consideration in planning research studies, is particularly significant in decision-making on electronic data collection. For EDC, it is important that the researcher assesses three types of resources: financial, time, and technological.

Financial Resources

Many studies that compare electronic and traditional data-collection methods focus on the issue of cost. Each study, however, defines cost in a different way, so each reaches a variety of conclusions about the relative costs of electronic and traditional collection methods. Despite these differences, researchers can make some generalizations regarding financial resources. Investigators must consider the technological cost of conducting electronic studies. For example, will the proposed study require the purchase of additional hardware or software? Note that EDC methods usually save personnel time spent on data entry and data cleaning, thereby increasing the speed of data analysis as well. Depending on the nature of the data collected, the personnel cost savings might outweigh the technological outlay.

Electronic surveys can offer considerable additional delivery and response savings in comparison to postal surveys (Raziano, Jayadevappa, Valenzuela, Weiner, & Lavizzo-Mourey, 2001; Schleyer & Forrest, 2000). In the Raziano et al. study, "the cost comparison showed that the average cost was $7.70 for the e-mail group, compared to $10.50 per response for the conventional group." In their study with dental professionals, Schleyer and Forrest's Web-based survey cost $1,916, while a comparable mail-based survey would have cost an estimated $3,092.

Time Resources

Researchers should evaluate time as a resource when selecting a data-collection method. As noted earlier, several studies have demonstrated more rapid data collection using electronic methods. Surveys deployed electronically can reach recipients and be returned more quickly than surveys relying on traditional mail (Raziano et al., 2001; Schleyer & Forrest, 2000). Also, electronic data usually do not require additional transcription and can often be imported directly into statistical-analysis software. Extensive time spent in data collection, especially in studies that require multiple responses from participants over time, is a burden for both participants and investigators.

Technological Resources

As with financial and time resources, technological expertise available to researchers must be weighed in data-collection decisions (Moloney,

Dietrich, Strickland, & Myerburg, 2003). Investigators lacking technological resources might find that time and money saved through EDC methods are offset by time and money spent recruiting and paying technological consultants.

Data output from commercial online services must often be manipulated into a useable format. Such transformation requires additional technological savvy. The researcher must be technologically competent in the service selected. For example, if the research will be using the communications of a virtual community, the researchers must be able to confidently use the communications software in their interactions with the virtual community (Moloney et al., 2003).

Conclusions

It would be nice if the EDC points discussed above were reducible to a tidy guideline detailing which data-collection method to use for which type of research question. There are too many scenarios in which such a prescriptive approach might overlook important considerations and guide researchers to inappropriate decisions. Researchers can instead use the population/data/resources decision factors to weigh the advantages and disadvantages and arrive at the optimal choice — likely a compromise — given the particular resources and priorities of the research setting.

The main population factors to consider in decisions about data-collection methods in relation to technological access are age, gender, socio-economic status, and race and ethnicity. The lower likelihood of some groups to be online — seniors, those without a college education, those with an annual household income below $30,000, and, to some extent, rural dwellers and non-white persons — can affect the representativeness of a sample participating in EDC. Online demographics are changing; access to these groups via e-mail and the Internet will likely increase.

Data-collection factors to consider in decision-making are sensitivity of the topic, time sensitivity, longitudinal data, and contextual data. Electronic methods consistently result in greater participant compliance than paper-based methods for longitudinal health data, especially when participants must report frequently. Participants tend to be more self-revelatory about sensitive subjects when reporting electronically. The question remains whether this holds true when the electronic method involves Internet use in a public place. For research relying on contextual data, however, traditional methods remain more effective; this might change as wireless technology improves.

Key factors to assess in data-collection decisions are financial, time, and technical resources. Electronic methods often result in savings of time

and money, compared to traditional methods, especially in the data-entry and data-cleaning phases of research and for surveys. Researchers must evaluate their technical infrastructure for available hardware, software, and expertise in order to make an accurate feasibility/cost comparison of data-collection methods.

Technology and demographics will continue to change and to affect data-collection methods and possibilities. By examining the key factors — population, nature of the data, and resources — the investigator will be able to make a data-collection decision based on the research endeavour.

References

Bell, P., Reddy, P., & Rainie, L. (2004). Rural areas and the Internet. *Pew Internet and American Life Project*. Retrieved July 17, 2004, from http://www.pewinternet.org/pdfs/PIP_Rural_Report.pdf

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90,* 125–144.

Chang, B. L. (2004). Internet interventions for community elders: Process and feasibility. *Western Journal of Nursing Research, 26*(4), 461–466.

Courtney, K. L., & Craven, C. K. (2004, May 21). *Electronic data collection: Advantages and disadvantages*. Paper presented at the American Nursing Informatics Association Spring Conference, New Orleans.

Cronk, B. C., & West, J. L. (2002). Personality research on the Internet: A comparison of Web-based and traditional instruments in take-home and in-class settings. *Behavior Research Methods, Instruments, and Computers, 34*(2), 177–180.

Fox, S. (2004). Older Americans and the Internet. *Pew Internet and American Life Project*. Retrieved July 17, 2004, from http://www.pewinternet.org/pdfs/PIP_Seniors_Online_2004.pdf

Greenspan, R. (2003, November 20). *Senior surfing surges.* Retrieved July 17, 2004, from http://www.clickz.com/stats/big_picture/demographics/article.php/3111871

Hinds, P. S., Chaves, D. E., & Cypess, S. M. (1992). Context as a source of meaning and understanding. *Qualitative Health Research, 2,* 61–74.

Jamison, R. N., Raymond, S. A., Levine, J. G., Slawsby, E. A., Nedeljkovic, S. S., & Katz, N. P. (2001). Electronic diaries for monitoring chronic pain: 1-year validation study. *Pain, 91,* 277–285.

Joinson, A. (1999). Social desirability, anonymity and Internet-based questionnaires. *Behavior Research Methods, Instruments, and Computers, 31*(3), 433–438.

Lenhart, A. (2005). Protecting teens online. *Pew Internet and American Life Project*. Retrieved June 4, 2005, from http://www.pewinternet.org/PPF/r/152/report_display.asp

Madden, M. (2003). America's online pursuits: The changing picture of who's online and what they do. *Pew Internet and American Life Project*. Retrieved July

16, 2004, from http://www.pewinternet.org/pdfs/PIP_Online_Pursuits_Final.PDF

Moloney, M. F., Dietrich, A. S., Strickland, O., & Myerburg, S. (2003). Using Internet discussion boards as virtual focus groups. *Advances in Nursing Science, 26*(4), 274–286.

O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments, and Computers, 35*(2), 217–226.

Palermo, T. M., Valenzuela, D., & Stork, P. P. (2004). A randomized trial of electronic versus paper pain diaries in children: Impact on compliance, accuracy and acceptability. *Pain, 107,* 213–219.

Pryor, J. H., Martin, M. T., Whitney, C. G., Turco, J. H., Baumgartner, Y. Y., & Zegans, M. E. (2002). Rapid response to a conjunctivitis outbreak: The use of technology to leverage information. *Journal of American College Health, 50*(6), 267–271.

Raziano, D. B., Jayadevappa, R., Valenzuela, D., Weiner, M., & Lavizzo-Mourey, R. (2001). E-mail versus conventional postal mail survey of geriatric chiefs. *Gerontologist, 41*(6), 799–804.

Rhodes, K., Lauderdale, D., He, T., Howes, D., & Levison, W. (2002). "Between me and the computer": Increased detection of intimate partner violence using a computer questionnaire. *Annals of Emergency Medicine, 40*(5), 476–484.

Schleyer, T. K. L., & Forrest, J. L. (2000). Methods for the design and administration of Web-based surveys. *Journal of the American Medical Informatics Association, 7,* 416–425.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Somenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*(5365), 867–873.

Authors' Note

Comments or queries may be directed to Karen L. Courtney, 324 Clark Hall, Columbia, Missouri 65211 USA. Telephone: 573-882-1625. Fax: 573-882-6158. E-mail: CourtneyKL@health.missouri.edu

---

Karen L. Courtney, RN, MSN, is National Library of Medicine Pre-doctoral Informatics Fellow, University of Missouri – Columbia, United States. Catherine K. Craven, MLS, is National Library of Medicine Informatics Research Fellow, University of Missouri – Columbia.